Relation between Speech Production and Speech Perception

Hiroya Fujisaki
Tokyo, Japan

1. Introduction

It is apparent that the mechanisms and processes of speech production and those of speech perception are essentially different. Functionally, speech is produced by successive execution of prestored motor programs, while speech is perceived by categorization of successive linguistic units and by recognition of their hierarchical structure. The acoustic characteristics of speech perceived by a speaker are not necessarily identical to those of speech which he/she produces, because of various contextual, idiosyncratic, and dialectical variations. On the other hand, it is quite natural to assume that the original message to be uttered by a speaker and the ultimate message to be received by the same person as listener would have the same form of linguistic representation in his mind. Moreover, many people tend to assume the existence of a further link between speech production and speech perception. In fact, several theories or models have been presented on their possible relations, such as the articulatory reference theory, the analysis-by-synthesis model (Halle and Stevens 1959), the motor theory of speech perception (Liberman et al. 1962), the auditory pattern model (Fant, 1967), the auditory-motor theory of speech production (Ladefoged et al., 1972), etc. Experimental evidences in support of these theories or models, however, have been rather scarce and indirect.

In this paper I will not try to review nor to criticize these theories or models, but will try to review some recent findings that will lead to a more concrete understanding of the possible links between speech production and speech perception. The works to be reviewed here can be classified into the following four areas:

1. Influence of speech production upon speech perception,
2. Influence of speech perception upon speech production,
3. Relation between production and perception in language learning and
4. Speech productions and perception by nonhuman vertebrates.

2. Influence of Speech Production upon Speech Perception

The human process of speech production is constrained by the physiological and physical properties of the mechanisms involved. Thus considerable

variations can be found in the realization of the same linguistic unit due to context, changes in speech rate, and other factors such as speaker idiosyncrasy. The process of speech perception has to separate these extraneous factors and extract the relevant linguistic information.

The use of contextual information in speech perception has been well known, but has recently been explicitly observed and measured quantitatively. For example, the presence of a vocalic context was found to modify the identification of the neighboring voiceless fricative consonant in such a way that the coarticulatory influence is compensated for in perception (Kunisaki and Fujisaki, 1977; Mann and Repp, 1980; Whalen, 1981). In the case of Japanese, anticipatory coarticulation by the vowel immediately following a fricative consonant is much stronger than perseveratory coarticulation by the preceding vowel, and the magnitude of perceptual compensation also reflects their difference.

Speech rate is another factor exerting influence on the acoustic characteristics of segments, especially on segmental duration. The existence of perceptual compensation against changes in speech rate has also been studied (Fujisaki, Nakamura and Imoto, 1973, Nooteboom, 1974, 1978: Johnson and Strange, 1982). The range of context used in perception has been found to extend beyond the immediately neighboring phonemes and syllables (Fujisaki, Nakamura and Imoto; 1973, Nooteboom and Cohen, 1975; Martin and Bunnell, 1981).

Another constraint imposed by the production mechanism is that of speaker idiosyncrasy or individual differences in the size and properties of speech organs due to age, sex and other factors. In a study using 3 male and 3 female speakers of American English, Fox (1982) reported a consistent perceptual structure difference of vowels among these subjects as listeners, and also found a high degree of correspondence between perceptual differences and articulatory differences among the subject. In view of the fact that speakers differing widely in the physical size and shape of their vocal tract (e.g. adults and children) can communicate reliably in ordinary situations, however, the above-mentioned influence from speech production to speech perception could not be of any significant magnitude. In fact, Paliwal et al. (1983), in a separate study using 10 speakers of British English, did not find any significant correlation between production and perception of vowels.

3. Influence of Speech Perception upon Speech Production

A syllable-timed language like Japanese imposes perception-based constraints on the temporal organization of speech production. As suggested by Lehiste (1977), isochrony is found to be a perceptual phenomenon, and is manifested as an approximate uniformity of perceived duration of syllables at least in the case of Japanese (Fujisaki and Higuchi, 1979). In a study of production and perception of dissyllabic words consisting of only two vowels, it was demonstrated that the perceived durations of the first and the

second vowels tended to be almost equal regardless of vowel combinations but that the articulatory onset of the second vowel, estimated by using a quantitative model of coarticulation varies rather widely depending on specific vowel combinations. There was a clear tendency that a slower articulatory transition is initiated earlier and *vice versa*, indicating that the apparent variability of onset of vowel articulation is the consequence of pre-programming to maintain the uniformity of perceived syllable duration. Even though the results could be explained both in terms of the chain model (i.e. assuming a closed-loop control) and in terms of the comb model (i.e. assuming an open-loop control, Kozhevnikov and Chistovich, 1965) it is more likely that such a perceptual isochrony should be achieved by an open-loop control based on well-prepared program for syllabic timing control (Allen and Tsukahara, 1974).

While the above-mentioned requirement on perceptual isochrony represents a form of perceptual constraint imposed on speech production, the existence of an immediate link from speech perception to speech production has also been suggested by an interesting experiment on perceptuo-motor adaptation (Cooper, 1974; Cooper and Nager, 1975). The result, however, was not replicated in a recent study by Summerfield et al. (1980). As I pointed out elsewhere (Fujisaki 1980), the selective adaptation paradigm is a useful tool, but the results have to be interpreted with caution since it creates a situation never to be encountered in natural speech communication.

4. Speech Production and Perception in Language Learning

The process of second-language learning provides another area of interest where the relation between production and perception of speech can be investigated. In a study on 72 native Spanish-speaking children learning English, both production and perception of voicing distinction in stop consonants (in terms of VOT) have been analyzed from the point of view of a) length of period of learning English, and b) the starting age (Williams, 1979). The results indicated that significant changes occur both in perception and production of voicing toward the pattern of monolingual English-speaking children selected as the control, and that the starting age had a significant effect in production, but not in perception.

The relation between production and perception of word accent was also investigated on 38 students of junior high school in Fukui learning the Tokyo dialect (Sugito and Fujisaki, 1980). The study revealed a strong correclation between the abibility of perception, as indicated by the accuracy of identification, and the ability of production, as indicated by the relative frequency of using correct accent types.

5. What do we learn from Experiments on Nonhuman Vertebrates?

A series of experiments have been reported on the perception of synthetic

speech sounds by non-human animals (chinchillas and macaques), and have indicated that these animals show human-like discriminability along certain consonantal continua (Kuhl and Miller, 1978; Kuhl, 1981, 1983), and also can be trained to identify certain vowels (Burdick and Miller, 1975). These results suggest that certain phonetic categories of human speech sounds are based on some psycho-acoustic properties that can be detected also by non-human animals, and provide a basis for an interesting hypothesis concerning the evolution of the human language. One might even say that speech perception takes place in the total absence of the ability of speech production. Similar arguments could also be made on speech production on the basis of the ability of mynahs and parrots which can produce excellent approximations to human sounds. The validity of these arguments will depend, not so much on the facts themselves, but rather on one's definition of speech production and speech perception.

In my opinion, the data on speech perception by animals and prelinguistic infants simply tell us that certain sounds of human languages are selected in such a way as to take advantage of some basic psychoacoustic properties of the sound continuum which can be generated by the human vocal apparatus. It is to be noted, however, that these experiments do not explain, at least up to the present, the origin of all existing categories of speech sounds, especially the categorization of the vowel continuum which is so much language-specific and can hardly be related to any basic psycho-acoustic properties.

References

Allen, G.L. and Tsukahara, N. (1974). Cerebro-cerebellar communication system. *Physiological Review,* **54**, 957-1006.

Bell-Berti, F., Raphael, L.J., Pisoni, D.B. and Sawush, J.R. (1979). Some relationships between speech production and perception. *Phonetica,* **36**, 373-383.

Burdick, C.K. and Miller, J.D. (1975). Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America* **58**, 415-427.

Cooper, W.E. (1974). Perceptuomotor adaptation to a speech feature. *Perception and Psychophysics,* **16**, 229-234.

Cooper, W.E. and Nager, R.M. (1975). Perceptuo-motor adaptation to speech: an analysis of bisyllabic utterances and a neural model. *Journal of the Acoustical Society of America,* **58**, 256-265.

Fox, R.A. (1982). Individual variation in the perception of vowels: Implications for a perception-production link. *Phonetica,* **39**, 1-22.

Fant, G. (1967). Auditory patterns of speech. In: *Models for the Perception of Speech and Visual Form.* (Wathen-Dunn, W., Ed.). 111-125, Cambridge, Mass.: The M.I.T. Press.

Fujisaki, H. (1980). Some remarks on recent issues in speech perception research. *Language and Speech,* **23**, 75-80.

Fujisaki, H. and Higuchi, N. (1979). Temporal organization of segmental features in Japanese disyllables. In: *Proceedings of the Ninth International Congress of Phonetic Sciences*, Copenhagen, 1979. Vol. II, 275-282.

Fujisaki, H., Nakamura, K. and Imoto, T. (1973). Auditory perception of duration of speech and non-speech stimuli. In: *Proceedings of the Symposium on Auditory Analysis and Perception of Speech*, Leningrad, 1973.

Halle, M. and Stevens, K.N. (1959). Analysis by Synthesis. In: *Proceedings of Seminar on Speech Compression and Processing, AFCRC, 1959, vol.* **2**.

Johnson, T.L. and Strange, W. (1982). Perceptual constancy of vowels in rapid speech. *Journal of the Acoustical Society of America,* **72**, 1761-1770.

Kozhevnikov, V.A. and Chistovich, L.A. (1965). *Speech: Articulation and Perception.* Moscow: Nauka.

Kuhl, P.K. and Miller, J.D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America,* **63**, 905-917.

Kuhl, P.K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conductive to the perception of speech-sound categories. *Journal of the Acoustical Society of America,* **70**, 340-349.

Kuhl, P.K. and Padden, D.M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America,* **73**, 1003-1010.

Kunisaki, O. and Fujisaki, H. (1977). On the influence of context upon perception of voiceless fricative consonants. In: *Proceedings of The Ninth International Congress on Acoustics,* Madrid, 1977. Vol. I, 476. Also in: *Annual Bulletin, RILP, University of Tokyo,* **11**, 85-91.

Ladefoged, P. DeClark, J.D., Lindau, M. and Papcun, G. (1972). An auditory-motor theory of speech production. *UCLA Working Papers in Phonetics* no. **22**, 48-75.

Lehiste, I. 1977). Isochrony reconsidered. *Journal of Phonetics,* **5**, 253-263.

Liberman, A.M., Cooper, F.S., Harris, K.S. and MacNeilage, P.F. (1962). A motor theory of speech perception. In: *Proceedings of the Speech Communication Seminar*, Stockholm, 1962. vol. **2**.

Mann, V.A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception and Psychophysics* **28**, 407-412.

Mann, V.A. and Repp, B.H. (1980). Influence of vocalic context on perception of the [š]-[s] distinction. *Perception and Psychophysics,* **28**, 213-218.

Mann, V.A. and Repp, B.H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America,* **69**, 548-558.

Martin, J.G. and Bunnell, H.T. (1981). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America,* **69**, 559-567.

Nooteboom, S.G. (1974). Some context effects on phonemic categorization of vowel duration. *IPO Annual Progress Report,* **9**, 47-54.

Nooteboom, S.G. (1978). Production and perception of vowel length in spoken sentences. *IP0 Manuscript* no. 335.

Nooteboom, S.G. and Cohen, A. (1975). Anticipation in speech production and its implications for perception. In: *Structure and Process in Speech Perception.* (Cohen, A. and Nooteboom, S.G. Eds.). 124-145. Berlin-Heidelberg-New York: Springer-Verlag.

Paliwal, K.K., Lindsay, D. and Ainsworth, W.A. (1983). Correlation between production and perception of English vowels. *Journal of Phonetics,* **11**, 77-83.

Repp, B.H. and Mann, V.A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America,* **69**, 1154-1163.

Repp, B.H. and Mann, V.A. (1982). Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America,* **71**, 1562-1567.

Sugito, M. and Fujisaki, H. (1980). Production and perception of word accent in speakers of the Fukui and the Osaka dialects. *Transactions of the Committee on Speech Research, the Acoustical Society of Japan,* **S80**, 221-228 (in Japanese).

Summerfield, Q., Bailey, P.J. and Erickson, D. (1980). A note on perceptuo-motor adaptation of speech. *Journal of Phonetics,* **8**, 491-499.

Whalen, D.H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of the Acoustical Society of America,* **69**, 275-282.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics,* **26**, 95-104.